State-specific protein–ligand complex structure prediction with a multiscale deep generative model

Year: 2022 | Citations: 131 | Authors: Zhuoran Qiao, Weili Nie, Arash Vahdat, Thomas F. Miller, Anima Anandkumar

---

Abstract

The binding complexes formed by proteins and small molecule ligands are ubiquitous and critical to life. Despite recent advancements in protein structure prediction, existing algorithms are so far unable to systematically predict the binding ligand structures along with their regulatory effects on protein folding. To address this discrepancy, we present NeuralPLexer, a computational approach that can directly predict protein–ligand complex structures solely using protein sequence and ligand molecular graph inputs. NeuralPLexer adopts a deep generative model to sample the three-dimensional structures of the binding complex and their conformational changes at an atomistic resolution. The model is based on a diffusion process that incorporates essential biophysical constraints and a multiscale geometric deep learning system to iteratively sample residue-level contact maps and all heavy-atom coordinates in a hierarchical manner. NeuralPLexer achieves state-of-the-art performance compared with all existing methods on benchmarks for both protein–ligand blind docking and flexible binding-site structure recovery. Moreover, owing to its specificity in sampling both ligand-free-state and ligand-bound-state ensembles, NeuralPLexer consistently outperforms AlphaFold2 in terms of global protein structure accuracy on both representative structure pairs with large conformational changes and recently determined ligand-binding proteins. NeuralPLexer predictions align with structure determination experiments for important targets in enzyme engineering and drug discovery, suggesting its potential for accelerating the design of functional proteins and small molecules at the proteome scale. Great advances in protein structure prediction have been made with recent deep learning-based methods, but proteins interact with their environment and can change shape drastically when binding to ligand molecules. To predict the 3D structure of these combined protein–ligand complexes, Qiao et al. developed a generative diffusion model with biophysical constraints and geometric deep learning.